

Face Recognition Using Scattering Convolutional Network

Shervin Minaee, Amirali Abdolrashidi and Yao Wang

ECE Department, NYU School of Engineering, USA
{shervin.minaee, abdolrashidi, yaowang}@nyu.edu

ABSTRACT

Face recognition has been an active research area in the past few decades. In general, face recognition can be very challenging due to variations in viewpoint, illumination, facial expression, etc. Therefore it is essential to extract features which are invariant to some or all of these variations. Here a new image representation, called scattering transform/network, has been used to extract features from faces. The scattering transform is a kind of convolutional network which provides a powerful multi-layer representation for signals. After extraction of scattering features, PCA is applied to reduce the dimensionality of the data and then a multi-class support vector machine is used to perform recognition. The proposed algorithm has been tested on three face datasets and achieved a very high recognition rate.

1. INTRODUCTION

Face recognition is currently one of the most popular tasks in computer vision, which has many applications in authentication, security, imaging technology, etc. Because of the variations in the images of the same person such as changes in facial expression, pose, viewpoint and lighting conditions, it could be quite challenging. A set of face images of one person with changes in facial expression and lighting are shown in Figure 1.



Fig. 1. Eight face images with different facial expressions

Various algorithms have been proposed for face recognition during the past years. Despite the huge progress in the development of algorithms that are robust to variations in illumination, pose and facial expression, there is still much room to explore for more reliable algorithms. There have been various approaches for the face recognition problem during past

years. Among those, Eigenface is one of popular approaches proposed in 1991 [1]. In this method, the face images are projected to a low-dimensional space using principal component analysis (PCA). These principal components which are derived from the training set are called Eigenfaces. In another work, known as Fisherfaces [2], instead of using PCA, the author used a class-specific dimensionality reduction algorithm known as linear discriminant analysis (LDA). In this approach, two covariance matrices are defined, one measures the inter-class variation and the other evaluates the intra-class variation, and the projection matrix is chosen as by maximizing the ratio of inter-class variation to intra-class variation. Because LDA takes the class label into account, it provides more discriminative power than PCA. Among more recent approaches, there have been a lot of works using sparse representation for face recognition. In [3], Wright used sparse coding for face recognition which achieved a high accuracy rate. In [4], Liu proposed an extended version of sparse-coding where he added a non-negative constraint on the sparse coefficients. There are also more recent works such as the one in [5], [6], which uses adaptive or robust sparse representation for face recognition. There are also many algorithms based on Bayesian framework for human face and pose recognition [7]-[8]. There have also been a lot of works using hand-crafted features for face recognition such as using SIFT, HOG, Gabor wavelet and curvelet [9]-[13]. In a more recent work, Guo [14] proposed a face recognition algorithm based on deep tree based structure, which uses a unified representation for multiple tasks, identity recognition, age estimation and facial expression.

In this paper, we propose an algorithm which uses a convolutional network called scattering transform/network which provides a multi-layer representation of the signal, and is invariant to translation, small deformation and rotation [15]. After derivation of scattering features, their dimensionality is reduced using PCA and fed into a multi-class SVM to perform classification. This algorithm has been tested on three face databases and achieved very high accuracy rate. The proposed algorithm is also fast enough to be used for mobile applications, using energy aware algorithms [16], [17].

The rest of the paper is organized as follows. Section 2 provides a description of the scattering features, which are used in this work and also a brief overview of PCA. Section

3 contains the explanation of the classification scheme. The experimental results and comparisons with other works are shown in Section 4 and the paper is concluded in Section 5.

2. FEATURES

Extracting good features is one of the most important steps in many of computer vision and object recognition tasks. As a result, a lot of research has been done in designing robust features for a variety of image classification tasks. Good features should be invariant to the transformations which do not change object class. Many image descriptors are proposed in the past two decades, including scale invariant feature transform (SIFT), histogram of oriented gradient (HOG), bag of words (BoW) [18]-[20].

Recently, unsupervised feature learning algorithms and deep convolutional neural networks have drawn a lot of attention and achieved state-of-the-art results in many computer vision problems [21]. They are shown to provide more abstract and discriminative features which suit better for object recognition task. In [15], a wavelet-based multi-layer representation is proposed, which is similar to deep convolutional network, where instead of learning the filters and representation, it uses predefined wavelets [22]. This algorithm has been successfully applied to digit recognition, texture classification and audio classification problems and achieved state-of-the-art results [15]. It has also been used for some biometric recognition tasks such as iris, fingerprint and palmprint recognition [23]-[25]. In this paper, the application of scattering transform for face recognition is explored. The details of scattering features and their derivation are presented in the following section.

2.1. Scattering Features

Scattering transformation is a multi-layer representation recently proposed by Stephane Mallat. The scattering transformation can be designed such that it is invariant to a group of transformations such as translation, rotation, etc. Here a translation-invariant version is used. It can be shown that the scattering coefficients of the first layer of the scattering network are similar to the SIFT descriptor, but the coefficients of the higher layers contain the high-frequency information lost in SIFT [15]. The scattering transform coefficients of each layer can be computed with a cascade of three operations: wavelet decompositions, complex modulus and local averaging.

For a given signal $f(x)$ we can derive its scattering representation as follows. The first scattering coefficient is just the averaged signal which can be obtained by convolving the signal with the averaging filter ϕ_J as $f * \phi_J$. Then the scattering coefficients of the first layer are obtained by applying wavelet transforms of different scales and orientations, taking the magnitude of wavelet coefficients and convolving it by the

averaging filter ϕ_J as shown below:

$$S_{1,J}(f(x)) = |f * \psi_{j_1, \lambda_1}| * \phi_J$$

where j_1 and λ_1 denote the scale and orientation respectively. By taking the magnitude of the wavelet, we can make these coefficients invariant to local translation. On the other hand, some of the high-frequency information of the signal will be lost by averaging. We can recover some of the lost information by convolving the term $|f * \psi_{j_1, \lambda_1}|$ by another set of wavelets at scale $j_2 < J$, taking the absolute value of wavelet followed by averaging as:

$$S_{2,J}(f(x)) = ||f * \psi_{j_1, \lambda_1}| * \psi_{j_2, \lambda_2}| * \phi_J$$

It is enough to only calculate the coefficients for $j_1 > j_2$, since $|f * \psi_{j_1, \lambda_1}| * \psi_{j_2, \lambda_2}$ is negligible for scales where $2^{j_1} \leq 2^{j_2}$. We can continue this procedure to obtain the coefficients of the k -th layer of the scattering network as:

$$S_{k,J}(f(x)) = ||f * \psi_{j_1, \lambda_1}| * \dots * \psi_{j_k, \lambda_k}| * \phi_J$$

$j_k < \dots < j_2 < j_1 < J, (\lambda_1, \dots, \lambda_k) \in \Gamma^k$

Figures 2 and 3 denote the transformed images of the first and second layers of scattering transform for a sample face image. These images are derived by applying a filter bank of 5 different scales and 6 orientations.

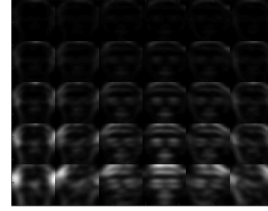


Fig. 2. Images from the first layer of scattering transform

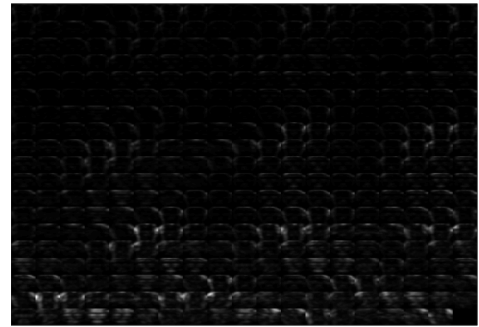


Fig. 3. Images from the second layer of scattering transform

As we can see, each image of the first layer is more sensitive to a specific orientation and scale. To derive scattering features in our experiment, the scattering-transformed images of all layers up to 2 are selected and the mean and variance

of these images are calculated and used as scattering features which results in a vector \mathbf{f}_s of size $\sum_{k=0}^2 2p^k \binom{J}{k}$. The scattering features can be extracted either locally or globally. Foreground segmentation schemes can also be used to detect important parts of the face, and extract features only around those regions [31]-[33].

2.2. Principal Component Analysis

Dimensionality reduction algorithms are an essential part of most of today's object recognition algorithms to reduce the feature dimension such that the discriminant power of features are kept. Principal component analysis (PCA) is a powerful algorithm used for dimensionality reduction [27]. Given a set of correlated variables, PCA transforms them to another domain such that the transformed variables are linearly uncorrelated. These linearly uncorrelated variables are called principal components.

Let us assume $\mathbf{x}_i \in \mathcal{R}^N$ denotes the features of the i -th image in the training set of n samples, and we want to find the orthonormal matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N]$ to project the images as $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$ such that the variance of projected \mathbf{y}_i 's is maximized. One can show that the optimum \mathbf{W} can be derived as: $\mathbf{W}^* = \text{argmax}_{\mathbf{W}} \|\mathbf{W}^T S_x \mathbf{W}\|$ such that $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, where S_x is the covariance of training images and is defined as $S_x = \sum_{i=1}^n (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T$. Then to reduce the dimension of the data to m , we can take the first m eigenvectors associated with the m largest eigenvalues of S_x and project the data on them.

3. RECOGNITION ALGORITHM: SUPPORT VECTOR MACHINE

After feature extraction, a classifier needs to be used to predict the label of each test image. Different machine learning algorithms can be used for classification. In this work, support vector machine (SVM) [28] is used to perform template matching. A brief overview of SVM in binary classification problem is provided here. Let us assume we are given a set of training data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ and asked to classify them into two classes where $x_i \in \mathbf{R}^d$ is the feature vector and $y_i \in \{-1, +1\}$ is the class label. For linearly separable classes, two classes can be separated with a hyperplane $w \cdot x + b = 0$. Among all possible hyperplanes which can separate two classes, one reasonable choice is the one with the maximum margin. The maximum margin hyperplane can be derived by the following optimization problem:

$$\begin{aligned} & \underset{w, b}{\text{minimize}} \quad \frac{1}{2} \|w\|^2 \\ & \text{subject to} \quad y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \dots, n. \end{aligned} \quad (1)$$

It turns out solving this problem in dual domain is simpler than primal domain and since this problem is con-

vex, the primal and dual solutions are the same. Then after derivation of Lagrange multipliers α_i in dual domain, we can derive the following hyperplane for classification: $f(x) = \text{sign}(\sum_{i=1}^n \alpha_i y_i x_i \cdot x + b)$, where α_i and b are calculated by the SVM learning algorithm. Surprisingly, after solving this optimization problem, most of the α_i 's are zero; therefore only the datapoints x_i with nonzero α_i are important in the final classifier. These points are called support-vectors. There is also a soft-margin version of SVM which allows for mislabeled examples by introducing a penalty term in the primal optimization problem with a penalty of C times the degree of misclassification [28].

To derive the nonlinear classifier, one can map the data from input space into a higher-dimensional feature space \mathcal{H} as: $x \rightarrow \phi(x)$, so that the classes are linearly separable in the feature space [29]. If we assume there exists a kernel function where $k(x, y) = \phi(x) \cdot \phi(y)$, then we can use the kernel trick to construct nonlinear SVM by replacing the inner product $x \cdot y$ with $k(x, y)$ which results in the following classifier:

$$f_n(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b\right) \quad (2)$$

To derive multi-class SVM for a set of data with M classes, we can train M binary classifiers which can discriminate each class against all other classes, and to choose the class which classifies the test sample with the greatest margin. In another approach, we can train a set of $\binom{M}{2}$ binary classifiers, any of which separates one class from another one and to choose the class that is selected by the most classifiers. Other schemes have also been proposed for multi-class SVM. For further detail and extensions to multi-class settings we refer the reader to [30].

4. EXPERIMENTAL RESULTS AND ANALYSIS

We have tested the proposed algorithm on three face databases, Yale Face Database, Georgia Tech Face Database and Extended Yale Face Database. We first discuss about the parameter values of our algorithm and then present the results of this algorithm on different databases.

4.1. Parameter Selection

For each image, the scattering transform is applied up to two layers using a set of filters with 5 scales and 6 orientations, resulting in 391 transformed images. The mean and variance of each scatter-transformed images are used as features, which results in scattering features of dimension 782. The scattering features are derived using the software implementation provided by Mallat's group [34]. Then PCA is applied to all features and the first K PCA features are used for recognition. Multi-class SVM is used for the classification. For SVM, we have used LIBSVM library [35], and linear kernel is used in our implementation.

4.2. Recognition Results on Three Databases

Yale Face Database: This section presents the results of the proposed algorithm on Yale Face Database. This database contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration. We have performed recognition using multi-class SVM. From each class, 6 images are used as training and the rest as test. We repeat the experiment for 5 different sets of training images and report the average accuracy here. Figure 4 demonstrates the recognition accuracy using different numbers of PCA features. For this case, the highest accuracy is achieved by using 200 PCA features which is around 93.1%.

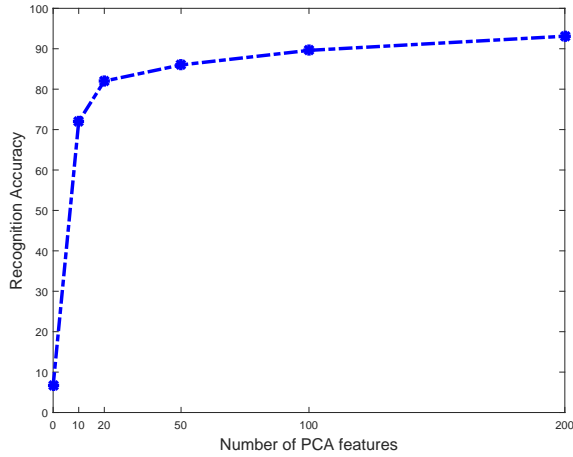


Fig. 4. Recognition accuracy for Yale Face Database

Georgia Tech Face Database: Georgia Tech Face database contains images of 50 people taken in two or three sessions. For each subject in the database, there are 15 color images with cluttered background with different facial expressions, lighting conditions and scales. For each person, 8 images are used as training and the rest as test. Figure 5 shows the recognition accuracy using different numbers of PCA features and multi-class SVM. For this case, by using first 100 PCA features, an accuracy rate of around 90% is achieved.

Extended Yale Face Database: This database contains the frontal face images of 38 individuals. We used the cropped images, which were taken under varying illumination conditions and each subject has around 64 images. We have used half of the images as the training and the rest as test. Figure 6 shows the recognition accuracy using different numbers of PCA features. For this case, by using the first 200 PCA features, an accuracy rate of around 85% will be achieved.

4.2.1. Comparison With Previous Works

Table 1 provides a comparison of the performance of the proposed scheme and that of 7 previous works on Yale face

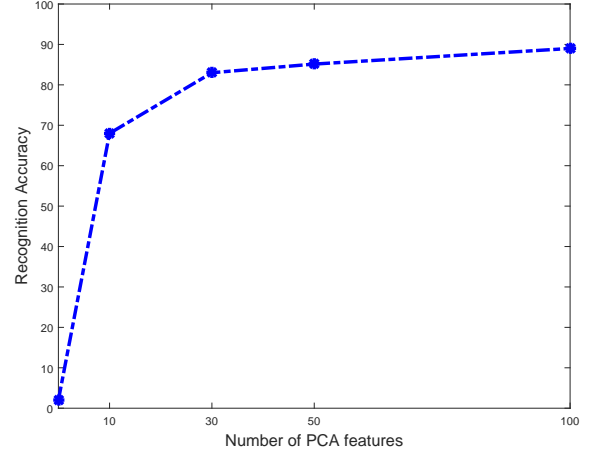


Fig. 5. Recognition accuracy for GaTech Face Database

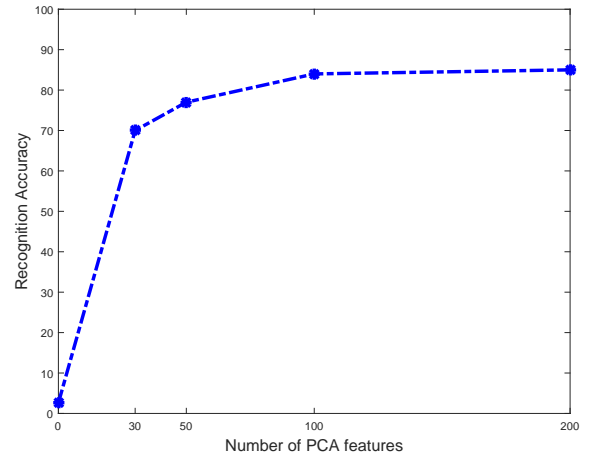


Fig. 6. Recognition accuracy for Extended Yale Database

database. The recognition accuracy for EigenFaces is taken from the reported values in [12]. We performed the experiment for 5 different sets of training images and reported the average accuracy below. As it can be seen, the proposed approach outperforms the previous schemes on this database.

5. CONCLUSION

This paper proposed a face recognition algorithm using scattering convolutional network. Scattering features are locally invariant and carry a great deal of high-frequency information, which are lost in other descriptors such as SIFT and HOG. After feature extraction, PCA is applied to reduce dimensionality. Then multi-class SVM algorithm is used to perform recognition. This algorithm has been tested on three well-known databases, and a high accuracy rate is achieved. The accuracy rate can be improved by using rotation-translation invariant scattering features, which is open for future research.

Table 1. Recognition accuracy comparison on Yale database

Face Recognition Method	Recognition Rate
EdgeMap [12]	74%
EigenFaces	76%
Sparse Representation Classifier [5]	82.34%
EigenFaces w/o first 3 PCA	84.7%
Correlation based classifier [5]	85.1%
Adaptive Sparse Representation [5]	88.06%
Curvelet + PCA + LDA [13]	92%
Proposed algorithm	93.1%

Acknowledgments

The authors would like to thank Mallat's research group for providing the software implementation of scattering transform. We would also like to thank the CSIE group at NTU for providing LIBSVM software, and also research groups at Yale and Gatech for providing the face databases.

6. REFERENCES

- [1] MA. Turk and AP. Pentland, "Face recognition using eigenfaces," IEEE conference on Computer Vision and Pattern Recognition, 1991.
- [2] PN Belhumeur, JP Hespanha and DJ. Kriegman, "Eigenfaces vs. fisher-faces: Recognition using class specific linear projection", Pattern Analysis and Machine Intelligence, IEEE Transactions on 19.7: 711-720, 1997.
- [3] J Wright, AY Yang, A Ganesh, SS. Sastry, and Y. Ma. "Robust face recognition via sparse representation", Pattern Analysis and Machine Intelligence, IEEE Transactions on 31, no. 2: 210-227, 2009.
- [4] Y. N. Liu, F. Wu, Z. H. Zhang, Y. T. Zhuang, and S. C. Yan, "Sparse representation using nonnegative curds and whey", In CVPR, 2010.
- [5] J Wang, C Lu, M Wang, P Li, S Yan and X. Hu, "Robust face recognition via adaptive sparse representation", IEEE Transactions on Cybernetics, 44(12), pp.2368-2378, 2014.
- [6] M Abavisani, M Joneidi, S Rezaeifar, SB Shokouhi, "A robust sparse representation based face recognition system for smartphones", IEEE Signal Processing in Medicine and Biology Symposium, 2015.
- [7] C Liu, H Wechsler, "A unified Bayesian framework for face recognition", International Conference on Image Processing, IEEE, 1998.
- [8] B Babagholami-Mohamadabadi, A Jourabloo, A Zarghami, S Kasaei, "A bayesian framework for sparse representation-based 3-d human pose estimation", IEEE Signal Processing Letters, 297-300, 2014.
- [9] C Geng, X Jiang, "Face recognition using sift features", In Image Processing (ICIP), 2009 16th IEEE International Conference on, pp. 3313-3316. IEEE, 2009.
- [10] O. Deniz, G Bueno, J Salido, F De la Torre, "Face recognition using histograms of oriented gradients", Pattern Recognition Letters 32, no. 12: 1598-1603, 2011.
- [11] C. Liu, H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition", Image processing, IEEE Transactions on 11, no. 4: 467-476, 2002.
- [12] Y. Gao, MKH. Leung, "Face recognition using line edge map", "Pattern Analysis and Machine Intelligence", IEEE Transactions on 24.6 : 764-779, 2002.
- [13] T Mandal, QMJ. Wu, Y Yuan, "Curvelet based face recognition via dimension reduction", Signal Processing, Elsevier, no. 12: 2345-2353, 2009.
- [14] R Guo, L Liu, W Wang, A Taalimi, C Zhang, and H Qi, "Deep tree-structured face: A unified representation for multi-task facial biometrics", IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2016.
- [15] J. Bruna, S. Mallat, "Classification with scattering operators," IEEE Conference on Computer Vision and Pattern Recognition, pp.1561-1566, 2011.
- [16] M Hosseini, A Fedorova, J Peters, and S Shirmohammadi, "Energy-aware adaptations in mobile 3D graphics", In Proceedings of the 20th ACM international conference on Multimedia, pp. 1017-1020, 2012.
- [17] M Hosseini, G Kurillo, SR Etesami, and J Yu, "Towards coordinated bandwidth adaptations for hundred-scale 3D tele-immersive systems", Multimedia Systems, 2016.
- [18] D. Lowe, "Distinctive image features from scale-invariant keypoints", International journal of computer vision 60.2: 91-110, 2004.
- [19] N Dalal, B Triggs, "Histograms of oriented gradients for human detection", IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Vol. 1. IEEE, 2005.
- [20] J Yang, YG Jiang, AG Hauptmann, CW. Ngo, "Evaluating bag-of-visual-words representations in scene classification", Proceedings of the international workshop on Workshop on multimedia information retrieval, pp. 197-206, ACM, 2007.
- [21] A. Krizhevsky, I. Sutskever, GE. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, 2012.
- [22] S. Mallat, "Group invariant scattering", Communications on Pure and Applied Mathematics 65, no. 10: 1331-1398, 2012.
- [23] S. Minaee, A. Abdolrashidi and Y. Wang, "Iris Recognition Using Scattering Transform and Textural Features", IEEE Signal Processing Workshop, 2015.
- [24] S. Minaee, Y. Wang, "Palmprint Recognition Using Deep Scattering Convolutional Network", arXiv preprint arXiv:1603.09027. 2016.
- [25] S Minaee and Y Wang, "Fingerprint Recognition Using Translation Invariant Scattering Network", IEEE Signal Processing in Medicine and Biology Symposium, 2015.
- [26] J. Bruna, S. Mallat, "Invariant scattering convolution networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, 35.8: 1872-1886, 2013.
- [27] H. Abdi and LJ. Williams, "Principal component analysis," Wiley Interdisciplinary Reviews: Computational Statistics 2.4: 433-459, 2010.
- [28] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning 20.3: 273-297, 1995.
- [29] B. Scholkopf, AJ. Smola, "Learning with kernels: Support vector machines, regularization, optimization, and beyond," MIT press, 2002.
- [30] J. Weston, C. Watkins, "Multi-class support vector machines," Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May, 1998.

- [31] S. Minaee, A. Abdolrashidi and Y. Wang, "Screen Content Image Segmentation Using Sparse-Smooth Decomposition", Asilomar Conference on Signals, Systems, and Computers, IEEE, 2015.
- [32] S. Minaee and Y. Wang, "Screen Content Image Segmentation Using Sparse Decomposition and Total Variation Minimization", International Conference on Image Processing, IEEE, 2016.
- [33] S. Minaee and Y. Wang, "Screen Content Image Segmentation Using Robust Regression and Sparse Decomposition", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2016.
- [34] <http://www.di.ens.fr/data/software/scatnet/>
- [35] CC. Chang, CJ. Lin, "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST) 2.3: 27, 2011.